⑲ 日本国特許庁（JP）

⑫ 公開特許公報（A）

⑪ 特許出願公開

昭59—23400

⑤Int. Cl.³ G 10 L 1/00

識別記号

庁内整理番号 7350—5D

㊸公開 昭和59年(1984)2月6日

発明の数 1

審査請求 未請求

（全 18 頁）

⑭音声認識装置

㉑特願 昭57—133573

㉒出願 昭57(1982)7月30日

⑫発明者 増子昭宣

深谷市幡羅町1丁目9番2号東京芝浦電気株式会社深谷工場内

⑪出願人 東京芝浦電気株式会社

川崎市幸区堀川町72番地

⑭代理人 弁理士 鈴江武彦 外2名

明 細 書

1. 発明の名称

音声認識装置

2. 特許請求の範囲

音声信号を複数のフィルタに通して複数の周波数帯域に分割し、これを同一タイミングで繰り返えしサンプリングすることにより音声信号の特徴を示すパターンを作るという操作によってそれぞれ登録モード時に得られる登録パターンデータと認識モード時に得られる入力パターンデータとを比較し、入力音声信号の内容を識別する音声認識装置に於いて、各サンプリング期間に於ける複数のサンプリングデータの振幅を正規化する振幅正規化手段と、各サンプリング期間に於ける複数のサンプリングデータの総和が一定しきい値以上のときを音声信号の始点とし、一定しきい値以下になったときを音声信号の終点とし、一定しきい値以上のデータとともに少なくともデータの始点以前のデータをも取り込むことが可能なデータ取り込み手段と、データの始点と終点間のデータの時間軸を正規化する時間軸正規化手段と、振幅及び時間軸の正規化された登録パターンデータ、入力パターンデータのどちらか一方を固定にし、他方のパターンデータをその始点データの格納されたアドレスを中心にして前後にずらすことにより固定にしたパターンデータの始点データに実質的に対応するようなデータが格納されているアドレスを検出するというようにして両パターンデータの比較を行なう比較手段とを具備した音声認識装置。

3. 発明の詳細な説明

〔発明の技術分野〕

本発明は、入力音声信号による命令、即ち話者の音声波から抽出された物理量の時系列を特徴パターンとしてとらえ、これをあらかじめ登録されたパターンと比較して音声信号による命令を認知する所謂、パターンマッチング法による音声認識装置に関する。

〔発明の技術的背景〕

一般に音声認識の方式は、音声信号から何ら

かの特徴を抽出した後得られる特徴（入力）パターンとあらかじめ登録されている登録パターンとの類似度を直接計算する方式と、前記音声信号から特徴を抽出した後にこれを音韻系列に置きかえ、これとあらかじめ登録されている単語辞書（パターン）とを比較して類似度を計算する方式の２つの方式に大別される。これら２つの方式のうち、後者は音韻単位の識別を行うために、単語数が多い場合の音声認識に優位である。しかし、単語数がさほど多くない場合には、前者によるパターンマッチング認識の方が一般に高い認識率が得られる。

認識される単語数が数１０程度の規模の前記パターンマッチングによる音声認識システムとしては、民生機器においては例えば、テレビジョン受像機を音声によって制御する場合が挙げられる。つまり、テレビジョン受像機の電源制御、音量制御、チャンネル切替等の制御を、あらかじめ音声認識装置に制御内容を表わす言葉の音声を登録しておき、応答装置には認識応答として音声を記憶させておき、音声命令と登録された制御内容とを照合して一致すると制御内容を認識したことを音声によって返答するとともに所定の制御をするような場合である。例えば、チャンネル切替制御において、１チャンネルを選ぶ場合、あらかじめ「１チャンネル」という音声を登録パターンとして記憶しておいたときに、音声命令を受信するマイクに向い「１チャンネル」という音声命令を下すと音声応答で「オーケー（ＯＫ）」と返答し、１チャンネルが選局される。

しかし、ここで問題となるのは、「１チャンネル」と音声命令を下した時に、これと音声が類似する「８チャンネル」という音声命令が制御パターン（登録パターン）として登録されている点である。即ち、「イチ」と「ハチ」の両者の音声は類似しており、「イチ」と「ハチ」とを誤まって音声認識するのをいかに防止するかが問題となる。これは、「イチ」という語と「ハチ」という語において、「チ」の発音部分の音声エネルギーが大きい為に、「イ」と「ハ」を区別するのが困難になることに起因する。一般に、一つの単語の中にアクセントをもつ音声があると、その部分に音声エネルギーが集中し、他の部分の音声情報の認識が困難となる。従って、音声認識に際しては、音声命令の強音以外の部分の情報を失うことなく特徴（入力）パターンと登録パターンとの比較をしなければならない。

また、話者が音声を発生する場合、同じ単語を発声しても、発声するたびに振幅が変化する。従って音声認識に際しては、振幅が変化しても同じ単語であれば常に同じパターンが得られるようにしなければならない。

また、制御内容を音声によって登録パターンとして登録する際の音声と、音声命令として発する音声の発生速度は必ずしも一致しない。このことは、ある単語を登録した後、その単語を再度同じように発声しても単語長は異なることを意味する。この為、入力パターンと登録パターン間の類似度を評価するに際しては、時間軸についても考慮しなければ誤認識がなされる。

第１図はパターンマッチング法に基づいた音声認識装置を示すブロック図である。発声による音圧振動をマイクロフォンで電気信号に変換し、更に前記音声の周波数分布を平坦化する機能を有する音声入力部１、この音声入力部１により得られる電気信号に変換された音声信号からその特徴を抽出する特徴抽出部２、この特徴抽出部２により抽出された特徴を記憶するとともにこれと入力パターンとの比較の演算処理を行ない音声による制御命令を判別する認識処理部３を有し、制御命令が認識されたことを音声により応答する音声応答部４が必要によっては付加される。この音声応答部４は、応答すべき言葉をパターンとして記憶してあるメモリ部４０１、第２のＩ／Ｏ（入出力）ポート４０２、制御部４０３、Ｄ／Ａ変換器４０４、ローパスフィルタ４０５を有しており、話者の音声指令が認知されたことをテレビジョン受像機４０６

等の被制御機器の音声回路から音声により応答する。

音声入力部1において、入力音声は、ワイヤレスマイク11によりFM波に変換した後FM受信機12で受信してプリアンプ13に入力する形態と、前記プリアンプ13前段に設けたマイクロフォン14によって入力する形態のいずれかによりシステムにとり入れられる。これらいずれの形態の場合においても、認識に必要な音声信号とそれ以外の音響信号との比であるSN比は、主としてマイクロフォンの指向性に左右されるのでマイクロフォン11，14は単一指向性のものを用いる。プリアンプ13に得られる電気信号に変換された音声信号は、単音節明瞭度を向上するため高音域をプリエンファシス回路15により強調する。

このようにして、得られる音声入力部1の出力は、特徴抽出部2に供給され、ここで入力及び登録パターンの形式に必要な特徴データの抽出処理がなされる。即ち、話者の音声波から時系列的に周波数をとらえ、音声を周波数分析しこれらのデータを一定時間間隔でサンプリングするとともに、サンプリングされたアナログデータをA/D変換器によりデジタル量に変換する。つまり、特徴抽出部2の入力端には16_1～16_{15}で示されるスイッチド・キャパシタ・バンドパスフィルタ（以下BPFと称する。）が接続されている。この16_1～16_{15}のBPFの中心周波数は印加されるクロックで決まり、その各々のフィルタ特性は6次のチェビシェフ特性で略−36dB/OCTの減衰特性を持つ。そして、前記BPF16_1～$_{15}$により、略200Hz～6.4KHzの帯域を1／3オクターブ間隔で15バンドに分離している。この15に分離されたバンドの帯域成分の音声信号を通過させる16_1～$_{15}$のBPFの夫々には、略20m Sec間隔で信号をサンプル・ホールドするサンプル・ホールド回路17_1～$_{15}$が接続されており、このサンプル・ホールド作用により到来する音声の特徴が抽出される。

このようにしてサンプル・ホールド回路17_1～$_{15}$に抽出された特徴データはアナログ量であるが、例えば8ビットのA/Dの変換器（アナログ−デジタル変換器）18によってデジタル量に変換される。このとき、前記サンプル・ホールド回路17_1～$_{15}$と前記A/D変換器18間の切換制御は、マルチプレクサ19によって行なわれる。従って、音声信号から抽出した、第2図に示す時間−周波数−レベルの特性をデジタル化した量が前記A/D変換器18に得られる。そして、このA/D変換器18で抽出された音声の特徴データは、第1のI/O（入出力）ポート20を介して認識処理部3に供給される。

この場合、I/Oポート20はプリエンファシス回路15の出力レベルがレベル検出器25に設定されるしきい値を越えたときを音声信号の始点とし、このときから8ビットA/D変換器18の出力をデータとして取り込む。そして、プリエンファシス回路15の出力レベルが上記しきい値以下になったときを音声信号の終点とし、その後のデータの取り込みを停止する。このしきい値は雑音信号によって越えられることがないように設定される必要があるが、これを満たす為にあまり高レベルに選ぶと、今度は認識に必要なデータが取り込まれなくなる危険性があるので、上記2つの条件をともに満足するような値に設定される。

認識処理部3は、制御内容、例えば受信するチャンネルの指定、電源のオン・オフの制御を音声によって指示する場合にその指令音声から抽出された音声の特徴を記憶させ登録するための登録パターンメモリ21、話者が希望する制御内容を発声した際にその指示音声の特徴を入力パターンとして一旦記憶するための入力パターンメモリ22、この入力パターンメモリ22の内容が前記登録パターンメモリ21に記憶された、いずれの登録パターンと類似するかの判定を行うためのプログラムを記憶するシステムプログラムメモリ23、このシステムプログラムの内容を実行するCPU（中央処理装置）24

からなる。そして、このCPU24は例えば、8ビットのマイクロプロセッサが用いられ、前記システムプログラムメモリ23は、2Kバイトの容量をもつROMで構成され、前記入力パターンメモリ22、登録パターンメモリ21は10Kバイトの容量をもつRAMによって構成される。この10KバイトのRAMのうち1.75Kバイトは入力パターンメモリ22として、略7.5Kバイトは登録パターンメモリ21として用いられる。

このような構成の認識処理部3に、前記特徴抽出部2で抽出されたデータが、入力パターンデータ、登録パターンデータとして送られる訳であるが、先ず登録パターンデータが送られる場合について述べる。

登録パターンデータが認識処理部3の登録パターンメモリ21に送られる場合は、前述の様に話者が希望する制御内容を何通りか発声により音声認識装置に登録する場合である。ここで、いま1チャンネルの選局を登録パターンメモリ21に制御内容として記憶させる場合についてみると、「1チャンネル」という音声の特徴データは前記A/D変換器18によってディジタルデータとして抽出される。そして、このデータは第1のI/Oポート20を介して登録パターンメモリ21に送られるが、このとき前記入力パターンメモリ22に次に示される行列式Aの形で一旦収納される。

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \\ \vdots & \vdots & & a_{ij} \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

m：サンプル回数。

n：フィルタの個数。（第1図の場合15）

a_{ij}：デジタル化されたサンプル値。

ここで、行列式の行数はサンプル回数、即ち、前記スイッチド・キャパシタ・バンドパスフィルタ16の出力が略20mSecの間隔のサンプルパルスに呼応してサンプルされる回数を示し、列数はBPF16の個数を示し、各成分はデジタル化された前記各BPFのサンプル値である。このようにして、抽出された話者の音声の特徴データは、未だ音声の振幅情報に対する正規化がなされていない。つまり話者のアクセントの位置或は強音によって弱音の情報が後退することに対する処理が行なわれていないので話者の音声の特徴を十分に表わしているとはいえない。そこで、前記行列式の各行の成分に対する加重を行う。即ち、前記Aで表わされる一旦、入力パターンメモリ22に収納されたデータに対してシステムプログラム23に記憶された次に示す演算をCPU24によって行ない演算結果の行列式αを前記登録パターンメモリ21に登録パターンとして格納する。

$$\alpha=\begin{pmatrix} \frac{a_{11}}{\sum_{j=1}^{n} a_{1j}} & \frac{a_{12}}{\sum_{j=1}^{n} a_{1j}} & \cdots & \frac{a_{1n}}{\sum_{j=1}^{n} a_{1j}} \\ \vdots & & & \vdots \\ \frac{a_{m1}}{\sum_{j=1}^{n} a_{mj}} & & & \frac{a_{mn}}{\sum_{j=1}^{n} a_{mj}} \end{pmatrix}=\begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ \vdots & & & \\ k_{m1} & k_{m2} & \cdots & k_{mn} \end{pmatrix}$$

このようにして、音声情報のうちの振幅情報は正規化される。この振幅の正規化は、話者が制御内容として発声する音声に対してすべてなされたうえで、前記登録パターンメモリ21にその内容（行列式）が記憶される。こうして、話者が発声により、前記登録パターンメモリ21に希望する制御内容を登録することで、音声認識装置に対する制御内容のセッテング は終了し、制御内容の数に等しい種類の登録パターン（α_1，α_2 ……α_n）が前記登録パターンメモリ21に記憶される。

上述のように、音声の特徴を示す行列式Aに対する振幅の正規化を行う演算は、前記システムプログラム23に記憶されたプログラム内容に応じてCPU24によって実行されるが、その実行内容を次に概式的に説明する。

即ち、前述の第1図中の第1のI/Oポート20、システムプログラムメモリ23、CPU24の動作は、次に示す第3図の機能動作に対応できる。

つまり、第3図中のラッチ回路301～15

（実際には入力パターンメモリ22に相当する。）には、前記行列式Aに相当するデータがラッチされ、ラッチされた内容は加算器31、及び乗算器32に夫々供給される。そして、この加算器31の出力は、レベル判定回路33と除算器341～16に供給される。前記加算器31は、前記行列式Aの各行成分の要素を加算し、

$$\sum_{j=1}^{n} a_{1j},\ \sum_{j=1}^{n} a_{2j} \cdots\cdots,\ \sum_{j=1}^{n} a_{mj}$$ を算出するが、

この夫々の総和値で前記ラッチ回路301～16にラッチされた行成分要素の各々が除算器341～16で除算される。ここで、除算器341～16の前段に乗算器321～16が設けられておりNなる乗算を行うが、これは前記除算結果を整数の形で評価するためのもので場合によっては省略し得る。また、前記の除算器341～16で除算され振幅が正規化されたデータは、バスラインを通して登録パターンとして、登録パターンメモリ21に収納される。

また、前記レベル判定器33には所定レベルの閾値が設定されており、前記加算器31の出力のレベルが設定された閾値以下の時は、前記ラッチ回路351～16のラッチされた内容をクリアし、それ以外の時は前記両ラッチ回路を制御しない。このように、ラッチ回路351～16に、前記加算器31の出力が一定値以上の時のみラッチ動作をさせることにより、検出する音声が小さい状態での雑音による誤動作が防止される。

上述の第3図の説明から判る様に、話者が希望する制御内容を登録パターンメモリ21に登録する過程において、振幅が正規化される前の特徴データは、一旦、RAMで構成される入力パターンメモリ22に記憶されこの後に振幅が正規化され、特徴パターンとして登録パターンメモリ21に記憶される。

次に、話者が登録した制御内容に対して、希望する制御内容を音声により指示した場合について述べる。

話者が、登録した制御内容のうち、希望する制御内容を発声し音声により指令をすると、音声の特徴データは登録パターンの時と同様に振幅が正規化され入力パターンメモリ22に記録される。ここで、話者が音声指令した内容に対し、その振幅に対する正規化を行なった入力パターンは次に行す行列式で示されるものとする。

$$F = \begin{pmatrix} f_{11}, f_{12}, \cdots\cdots f_{1n} \\ f_{21} \quad \vdots \\ \vdots \quad \vdots \\ f_{m1}, f_{m2} \cdots\cdots f_{mn} \end{pmatrix}$$

この振幅が正規化され入力パターンメモリ22に記憶される入力パターンFは、既に制御内容として登録パターンメモリ21に登録されている登録パターンとの参照が行われる。この参照動作による両パターン間の類似度の演算処理により、類似度が一番近いパターンに対応する制御内容を話者が指令した制御内容であると判定する。

このような入力パターンと登録パターンの両パターン間の類似度は、次に示されるパターン間の距離Dを計算することにより判別される。即ち、前記振幅が正規化された登録パターンKと入力パターンFと各成分kij，fijの差の絶対値をとることにより得られる行列式を両パターン間の距離を表わす行列式距離パターンDと定義し、この行列式Dの各成分の総和値によって類似度を算出する。このことを更に述べると、前記距離パターンDは次式で示され、かつ類似度dは次のように示される。

$$D = \begin{pmatrix} |k_{11}-f_{11}|, |k_{12}-f_{12}|, \cdots\cdots, |k_{1n}-f_{1n}| \\ |k_{21}-f_{21}|, \\ \vdots \\ |k_{m1}-f_{m1}|, |k_{m2}-f_{m2}|, \cdots\cdots, |k_{mn}-f_{mn}| \end{pmatrix}$$

$$= \begin{pmatrix} r_{11}, r_{12} \cdots\cdots, r_{1n} \\ r_{12} \quad \vdots \\ \vdots \quad \vdots \\ r_{m1} \quad r_{m2} \cdots\cdots, r_{mn} \end{pmatrix}$$

$$d = \sum_{i=1}^{m} \sum_{j=1}^{n} r_{ij}$$

上記、類似度dの計算は全登録パターン、いいかえると全制御内容を表わすパターンに対して行われ、類似度dの値が最っとも小さいパターンを話者が音声によって指令したパターンであると判定する。このようにして音声認識が行われるが、上述のように音声の振幅に対する正規化を行うことで誤認識率は著しく低減される。話者の発声に対する音声認識はこうして、登録パターンと入力パターンの類似度が、前記システムプログラムメモリ23に設定された類似度算出プログラムによって指示される演算が前記CPU24で実行されることにより算出され、音声認識による機器の制御が可能となる。

上述した音声のパターン・マッチング法による音声認識では、振幅が正規化されることで単語中の強音部分に比較して弱音部分の情報が小さい点及び第4図に示すように同じ単語でも発声のたびに振幅が変動しやすい点に起因する音声の誤認識は低減される。なお、第4図は例えば第2図の時刻(t_1)における周波数スペクトルを示すもので、実線及び破線はそれぞれ同じ単語を大きく発声した場合及び小さく発声した場合を示す。

ところで、前述の如く話者が同一単語を発声してもその発声時間が常に一致するとは限らない。この問題を解決するには時間軸についても正規化を行なうことが必要であり、次にこの時間軸の正規化について説明する。時間軸の正規化は、話者の発音単語の発音開始時刻と発音終了時刻との間にかかる時間を、常に一定の定数nで分割することによりなされる。つまり、話者がある単語を発声するにある時は時間T_1かかり、またあるときには時間T_2を要した場合、それぞれの場合、特徴抽出のためのサンプル時間間隔を$\triangle T_1 = \frac{T_1}{n}$, $\triangle T_2 = \frac{T_2}{n}$ とすることで解決される。このことは、時間軸のずれに呼応して音声の特徴が生起する時刻がずれるという現象に根拠をおく。従って、話者の発声の開始時刻と終了時刻は極力正確に検知する必要がある。前述のように、入力パターン、登録パターンのいずれの場合においても話者の音声の特徴の抽出は、BPF $16_1 \sim 16_{15}$、サンプル・ホールド回路 $17_1 \sim 17_{15}$ の両者に依存するが、両回路はいずれもその動作に時定数的な要素をもつ。とりわけ、サンプル・ホールド回路のピーク検波方式は話者の発声の終了時刻の検出を正しく行うのに大きく左右する。従って、特徴抽出部2を構成するサンプル・ホールド回路におけるピーク検波方式、及びサンプリングのタイミングは話者の発声長を正確にとらえた上で時間軸の正規化を行うのに重要な点となる。

次に、時間軸の補正を適格にするに適した特徴抽出部2の他の例について説明する。

一般に話者がある単音を(第5図ⓐに示す音声波形)発声すると、前記BPF $16_1 \sim 16_{15}$ の出力には第5図ⓑに示すように、ピーク値間のピッチがPの波紋が得られる。このピッチPは、例えば「ア」という単音を発声した場合には約8 m Secであるが、普通の音声ではこのピッチは5～15 m Sec以内に入いる。このようなピッチPを有する第5図ⓑに示されるBPF $16_1 \sim 16_{15}$ の出力は、夫々第5図ⓒに示される様にピーク検波されるわけであるが、検波するときの時定数によっては第5図ⓓ, ⓔに示されるように発声の終了時刻を誤まって検出する。即ち、ピーク検波によるリップルを少なくするために時定数を大きくすると、検波出力は第5図ⓓで判るように、時刻t_1で実際には発声が終了しているにも拘らず、時刻t_2まで音声が継続していると認識する。また、これに対して時定数を小さくした場合には、検波波形にリップルが生じて正確な特徴パターン抽出が望めない。このことは、時間軸の正規化と特徴パターンの抽出に影響を与え誤った音声認識を行う原因ともなる。

そこで、近時ピッチ周期より長い周期でピーク値検出を行う方法が考えられている。以下この方法について図面を参照して説明する。

第6図は、第1図に示した特徴抽出部3の他の例を示す回路ブック線図であり、入力端子P_1に音声入力部1(図示せず。)からの音声信号

が BPF 41 ～ n に供給される。そして、この BPF 41 ～ n の各々の出力はダイオード D1 ～ n と、ピーク検出機能を有するサンプル・ホールド回路 42 ～ n を構成する MOS トランジスタ Q1 ～ n 及びピーク値をホールドするコンデンサ C1 ～ n によってピーク検波される。ピーク検波によって検出されたピーク値、即ち、音声の振幅データは前記コンデンサ C1 ～ n に保持され、これらの振幅データは 2 進 - 10 進デコーダ 43 と MOS トランジスタ Q'1 ～ n よりなるマルチプレクサ 44 を介して A/D 変換器 45 に供給される。ここで前記 MOS トランジスタ Q1 ～ n がオンのときは前記マルチプレクサ 44 を構成する MOS トランジスタ Q'1 ～ n はオフの状態であり、一方のトランジスタ群がオンのときは他方のトランジスタ群がオフとなる様に制御されている。このため、前記 MOS トランジスタ Q1 ～ n がオンのときコンデンサ C1 ～ n に保持された音声の振幅データは、前記 MOS トランジスタ Q1 ～ n がオフのときに MOS トランジスタ Q'1 ～ n を介して A/D 変換器 45 に供給されデジタル量に変換される。前記ピーク値のサンプリングは、前述したピッチ P の時間より長い時間 T で行なわれ、時間 T だけピーク値が保持されるとその後、トランジスタ T1 ～ n、抵抗 R1 ～ n、R'1 ～ n によって構成されるリセット回路 46 によって前記コンデンサ C1 ～ n の充電電荷は放電される。この放電時間後、再びピーク値の検出が開始されこれを話者の発声の終了までくり返す。第 7 図を用いてこのことを説明すると、第 7 図ⓐは BPF 41 ～ n のうちの 1 つの出力を示し、同図ⓑに示す時間 T のサンプリングパルスで音声のピーク値が検出されるとともにピーク値が保持され、同図ⓒに示すリセットパルスでコンデンサ C1 ～ n の充電電荷は放電されるので、A/D 変換器 45 の入力には同図ⓓに示す波形が入力される。第 7 図で判るように音声のピーク値は、前述のピッチ P よりも長い時間 T だけ保持され、しかも放電時はリセットパルス期間なので、放電による誤まった検波出力の振幅データを A/D 変換器 45 に送ることもない。

次に前記の T なる時間、ピーク値をサンプル保持するためのサンプリングパルスを発生させる手段及びリセットパルスを発生させる手段について第 6，8，9 図を用いて説明する。前記コンデンサ C1 ～ n に音声のピーク値をサンプル保持するためサンプリングパルスは、分周器 47 とナンドゲート 48 によって得られる。

即ち、分周器 48 のクロック端子 CK には、第 8 図の CK で示されるクロックパルスが印加され、これを分周して Q_0，Q_1 に示される出力をナンドゲート 48 に印加することにより第 8 図中ⓗで示すサンプリングパルスが得られる。このサンプリングパルスが前記 MOS トランジスタ Q1 ～ n の導通を制御することは前述の通りである。また、第 1 図のモノマルチ 49 は前記サンプリングパルスⓐの立ちさがりを検出してパルス（第 8 図ⓑ）を発生しフリップフロップ 50 の出力を反転する（第 8 図ⓒ）。すると、ナンドゲート 51、インバータ 52 を介して第 9 図に示すクロックパルス CK' が m ビットカウンタ 53 に印加されこのクロックパルス CK' をカウントし始め前記マルチプレクサ 44 を構成する 2 進 - 10 進デコーダを順次切替え、全てのスキャンが終わると前記 m ビットカウンタ 53 の出力 Q がインバータ 54 を介して前記フリップフロップ 50 にリセットパルスとして供給され、フリップフロップ 50 の状態が再び反転する。そして、これと同時に第 2 のモノマルチ 55 が前記トランジスタ T1 ～ n を導通させコンデンサ C1 ～ n の充電電荷を放電させるリセットパルス（第 8 図，第 9 図ⓓ，第 7 図ではⓒに相当する。）を発生する。

尚、分周器 47 に接続された、イニシャライズ回路 57 は、電源投入時に前記分周器 47 をリセットするためのもので(R)は抵抗，(D)はダイオード，(C)はコンデンサである。

また、前記 A/D 変換器 45 へのデータの読み込みのタイミングは次のようにして第 9 図ⓕに

示すパルスを発生することにより行なわれる。前述のように、サンプリングパルス（第8図ⓐ）の立ち下がりで、第1のモノマルチ49はパルス（第8，9図ⓑ）を発生する。このパルスによりフリップフロップ50の状態は反転し（第8，9図ⓒ）、mビットカウンタ53にはクロックパルスCK'（第9図ⓔ）が印加される。このクロックパルス（第9図ⓔ）の立ち下がりは第3のモノマルチ56で検出され、この第3のモノマルチ56の出力には第9図ⓕで示されるパルスが発生される。そして、このパルスが前記A/D変換器45のデータ取り込みタイミングパルスとして用いられる。

このようにして、近時、単音発声時にみられる前述のピッチPより大きい時間Tを音声の特徴抽出のためのサンプル時間とし、ピーク検波時においてリップルによる音声認識時における誤った特徴抽出を防止するようにしている。また、話者の発声終了時刻の判定に際しても、その誤差範囲を略前記ピッチ長Pよりも少ない範囲とすることができるので、時間軸に対する正規化を行うにあたり誤認識を低減できる。いいかえると、話者が同一の単語を発声するに要する時間を発声のたびに異ならせたとしても、このことによる音声の認識を低減することができる。

〔背景技術の問題点〕

しかしながら上記構成の場合、次のような問題がある。すなわち、話者が同じ制御内容の言葉を発生したとしても、話者とワイヤレスマイク11やマイクロフォン14との距離、発声の強さ等によりワイヤレスマイク11，マイクロフォン14に入力される音声信号の第10図に示すように振幅レベルが変化する。今、実線で示すパターンが登録時のものとし、破線で示すパターンが命令時のものとすれば、登録パターンメモリ21には期間T_1のデータが取り込まれるのに対し、入力パターンメモリ22には期間T_2のデータしか取り込まれない。このように入力音声信号のレベルが異なると、登録パターンメモリ21に記憶されているデータと入力パターンメモリ22に記憶されているデータとが違ってくる為に、実際には同じ制御内容の言葉であるにもかかわらずそれと認識されない誤認識が発生する。

〔発明の目的〕

この発明は上記の事情に対処すべくなされたもので、入力音声信号のレベルが異なる為に登録パターンメモリと入力パターンメモリとに記憶されるデータが異なってしまい、誤まった認識動作が行なわれてしまうことを防止し得る音声認識装置を提供することを目的とする。

〔発明の概要〕

この発明は始点と終点間のデータだけでなく、少なくとも始点以前のデータも取り込むようにし、登録パターンデータと入力パターンデータのどちらか一方を固定にし、他方をそのスタートアドレスを中心に前後にずらすことにより、固定にしたパターンデータのスタートアドレスのデータに実質的に対応したデータが格納されているアドレスを検出し、これに基づいて両パターンデータの距離を計算して認識処理を行なうように構成したものである。

〔発明の実施例〕

以下、図面を参照してこの発明の一実施例を詳細に説明する。第11図は一実施例の回路図で、先の第1図及び第6図と同一部には同一符号を付して説明する。サンプルホールド回路17_1～17_5によってピーク検出されたデータはマルチプレクサ19で切り換えられ、8ビットA/D変換器18でデジタルデータに変換される。このデジタルデータはラッチ回路51に一時蓄えられる。最大値検出回路52はラッチ回路51にラッチされたサンプルホールド回路17_1～17_5の出力データの中の最大値を検出するとともに、全ラッチデータを加算する。ラッチ回路51のラッチデータは割算回路53に供給され、最大値検出回路52で検出された最大値を用いて割算される。この動作は前述したような振幅の正規化に相当するものであり、この振幅の正規化されたデータは入力パターンメモリ54に

記憶される。なお、最大値を用いて正規化することは本件出願人が先の特願昭55－88019号にて出願したものであり、先の第3図で説明した全ラッチデータの加算値で割算する構成に比べ、認識率を高めることができる。

最大値検出回路52のもう1つの出力、つまり全ラッチデータの加算出力はしきい値検出回路55に供給される。このしきい値検出回路55は予じめ設定された一定レベルのしきい値V_Tにより、最大値検出回路52から出力される加算出力がしきい値V_T以上かしきい値V_T以下かを識別するの為の信号を出力する。この識別信号は制御回路56に供給される。この制御回路56は例えばマイクロコンピュータから成り、先の第1図に示すようなシステムプログラムメモリ23，CPU24等を有する。制御回路56は上記加算出力が第12図に示す如く、上記しきい値V_Tを越えてから一定時間経過しても今だしきい値V_T以上であるときは、入力信号が雑音信号ではなく話者の音声信号であると判断する。そして、しきい値V_Tを越えたときのデータが記憶されている入力パターンメモリ54のアドレスを音声信号の始点のデータを記憶するスタートアドレスA_Sとしてスタートアドレスメモリ57に記憶する。また、制御回路56は上記加算出力がしきい値V_Tより小さくなってから一定時間経過しても今だしきい値V_T以下であるときは、発声の終わりと判断する。そして、しきい値V_Tより小さくなったときのデータが記憶されている入力パターンメモリ54のアドレスを音声信号の終点のデータを記憶するエンドアドレスA_Eとしてエンドアドレスメモリ58に記憶する。一定時間しきい値以下であるか否かをチェックするのは発生の途中なのか、単語の区切りなのかを区別する為である。一般に発生の途中でしきい値以下となるのは150 msec 以下であり、それ以上は単語の終わりとみなす。なお、入力パターンメモリ54の書き込みアドレス指定は制御回路56によって駆動されるアドレス発生回路59から出力される書き込みアドレス指定データによってなされる。制御回路56はこの他にも、サンプルホールド回路17_1～17_{15}のサンプリング動作用のクロックパルスやマルチプレクサの切り換えタイミングを指定するタイミングパルス、8ビットA/D変換器18の変換タイミング、ラッチ回路51のラッチタイミング等を指定するタイミングパルスを出力する。

第13図は入力パターンメモリ54の記憶状況を示すもので、図示の如くメモリ54にはしきい値V_T以上のデータの他にしきい値V_T以下のデータも数アドレス分にわたって記憶されている。この点に関し、先の第1図の装置はしきい値V_Tを越えた部分のデータだけを入力パターンメモリ22に記憶するものであった。

今、話者が発生した各種制御内容を示す音声信号を登録する登録モードであるとすると、入力パターンメモリ54に記憶されているデータはアドレス発生回路59から出力される読み出しアドレス指定データに従って読み出され、時間軸正規化回路60にて正規化され、登録パターンメモリ61に記憶される。一方、話者が所望の制御内容を指令する認識モードであるときは、入力パターンメモリ54の記憶データと登録パターンメモリ61の記憶データとの比較がパターンデータ比較回路62によってなされる。この場合、入力パターンメモリ54の入力パターンデータは時間軸正規化回路60によって時間軸の正規化がなされた後にパターンデータ比較回路62に供給される。なお、時間軸正規化回路60はスタートアドレスからエンドアドレスまでのデータを例えばM等分して得たM個のデータを正規化データとし、これに始点の前及び終点の後のデータを数アドレス分付加したものを出力し、登録モードであれば この出力は登録パターンメモリ61に記憶され、入力モードであれば、パターンデータ比較回路62に供給される。

パターンデータ比較回路62は例えば登録パターンデータを固定にし、入力パターンデータをそのスタートアドレスを中心にして前後にず

らすことにより、入力パターンの各アドレスに於いて登録パターンデータの始点データに実質的に対応するようなデータが格納されているアドレスを検出し、両パターンデータの比較を行なう。このように登録パターンデータの始点に実質に対応するアドレスを検出することにより、音声信号の入力レベルが変化することによって取り込まれるデータが同じ内容の音声信号によるものでありながら異なってしまうことに起因する誤認識を低減させることができる。

これに対し、従来は予じめ設定された一定のしきい値によって決まる始点と終点間のデータで距離を求め、その距離が所定のレベル以下であれば、認識条件を満たしたとして、認識内容に応じた処理及び表示を行なうようにしていた。したがって、音声信号の入力レベルが変化すると、同じ内容の音声信号でありながら取り込まれるデータの内容が異なってしまい、認識不能あるいは誤認識が行なわれることがあったわけである。

第14図は第11図の装置の動作を示すフローチャートである。ステップS_1によって認識モードか否かが判別され、認識モードでなければ、ステップS_2によって登録キーが操作されていることを検出し、ループL_1で示される登録モードに入る。ステップS_3によって振幅の正規化がなされる。ステップS_4は音声信号の始点を検出するステップである。このステップS_4の処理には例えば4ビットの循環レジスタが用いられる。すなわち、音声信号の入力レベルがしきい値V_T以下であれば、循環レジスタの内容は"4"で変化しないが、しきい値以上になると各サンプリング期間毎に"1"ずつ減らされる。そして、循環レジスタの内容が"0"になったとき、入力信号が雑音信号ではなく音声信号であると判断して、しきい値V_Tを越えた最初のデータが格納されるいるアドレスをスタートアドレスとして記憶する（ステップS_5）。ステップS_6はエンドアドレスを記憶するステップであり、ステップS_4と同様に4ビット循環レジスタを用いて処理される。但し、この場合、音声信号の入力レベルがしきい値以下になると、循環レジスタの内容が変化し、N＝0になったとき、発声の終わりと判断して、例えばしきい値V_T以下になったときの最初のデータが格納されているアドレスをエンドアドレスとして記憶する（ステップS_7）。このように入力パターンメモリ54に対するデータの記憶とスタート及びエンドアドレスの記憶が並列して行なわれ、これが終了すると、ステップS_8で時間軸の正規化がなされ、ステップS_9で登録パターンメモリ61へのデータの記憶がなされる。

ループL_2は認識モードであり、ステップS_3'～S_8'はそれぞれ先のステップS_3～S_8に対応する。ステップS_8'で時間軸の正規化が終了すると、ステップS_{10}に移る。このステップS_{10}は前述のパターデータ比較回路62の動作に対応するものであり、その動作の一例を示している。すなわち、ステップS_{101}で登録パターンメモリ61に記憶されている複数の登録パターンデータそれぞれのスタートアドレスのデータと入力パターンデータのスタートアドレスのデータとの距離が計算される。この計算結果に基づいて、ステップS_{102}距離が最も小さい登録パターンデータが最適パターンデータとして選択される。そして、ステップS_{103}で選択された登録パターンデータを固定にし、入力パターンデータをそのスタートアドレスを中心に前後に1アドレス分ずらし距離を計算する。この1アドレス分ずらす操作によって全体の距離が小さくなった方向に対して入力パターンデータをずらし、登録パターンデータのスタートアドレスデータとこのスタートアドレスに対応した入力パターンデータのアドレスのデータとの距離を計算し、最も小さくなったアドレスを実質的に登録パターンデータのスタートアドレスに対応するものとし、全体の距離を計算する。入力パターンデータのアドレスを前方向に±1したのであれば、エンドアドレス側を同方向に∓1し、距離を計算する上での登録パターンメモリ

と入力パターンメモリとのアドレス数をそろえる。そして、全体の距離が所定レベルより小さければ、ステップS_{104}にて認識条件を満たしたと判断して、ステップS_{105}にて制御内容に基づいた機器の制御やその制御内容の表示を行なう。認識条件を満たさなければ、ステップS_1に移って、再び話者に希望の制御内容を発声させる。

以上詳述したように動作させる場合、例えばステップS_6，S_8で循環レジスタの内容が"0"になったときデータの取り込みを終えるようにすることができるので、無駄なアドレス制御等の不要な演算を無くし、実行時間を短縮してデータの転送効率を高めることができる。また、ステップ103に於いて、入力パターンデータを±1だけずらし、この後は全体の距離が小さくなった方向にだけ入力パターンデータをずらすことにより最適ポイントを見つけるようにしたので、前後にそれぞれ数アドレス分ずらして最適ポイントを見つける場合に比べ処理時間を短縮できる。

なお、この発明は入力パターンデータを固定にして登録パターンデータを前後にずらすようにしてもよい。また、始点と終点との間のデータ以外のデータの取り込みは登録パターンデータ，入力パターンデータのどちらか一方に対してのみ行なうようにしてもよい。

〔発明の効果〕

このようにこの発明によれば、同じ内容の音声信号であっても取り込むデータの内容が異なってしまって誤認識が生じてしまういうことを無くし得る音声認識装置を提供することができる。

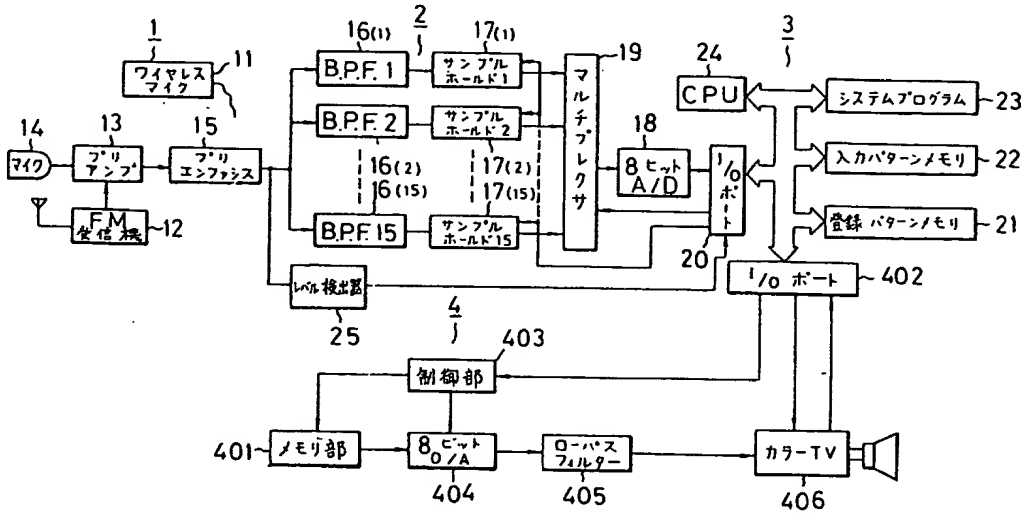
4. 図面の簡単な説明

第1図は音声認識装置として現在考えられているものを示す回路図、第2図，第3図，第4図はそれぞれ音声認識装置として現在考えられているものの説明に供する時間－周波数－振幅レベル特性図、回路図、周波数スペクトル図、第5図は音声波の検波特性を説明するに供する信号波形図、第6図は音声認識装置として現在考えられているものの他の例を示す回路図、第7図は第6図の動作を説明するに供する信号波形図、第8図及び第9図は第6図の動作を説明するに供するタイミングチャート、第10図は第1図の装置の欠点を説明する為の図、第11図はこの発明に係る音声認識装置の一実施例を示す回路図、第12図は一定レベルのしきい値によって設定される音声信号の始点及び終点を説明する為の図、第13図は入力パターンメモリの記憶状況を説明する為の図、第14図は第11図の装置の動作を説明する為の図である。

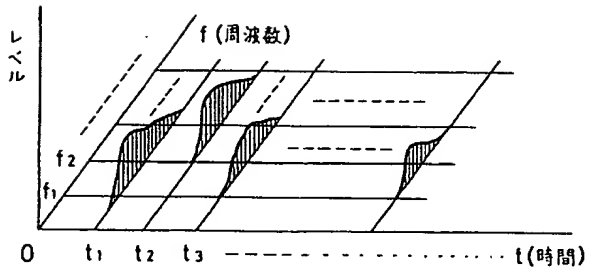
11…ワイヤレスマイク、12…FM受信機、13…プリアンプ、14…マイク、15…プリエンファシス回路、16_1～16_{15}…バンドパスフィルタ、D_1～D_{15}…ダイオード、17_1～17_{15}…サンプルホールド回路、18…8ビットA/D変換器、19…マルチプレクサ、51…ラッチ回路、52…最大値検出回路、53…制算回路、54…入力パターンメモリ、55…しきい値検出回路、56…制御回路、57…スタートアドレスメモリ、58…エンドアドレスメモリ、59…アドレス発生回路、60…時間軸正規化回路、61…登録パターンメモリ、62…パターンデータ比較回路。

出願人代理人　弁理士　鈴　江　武　彦

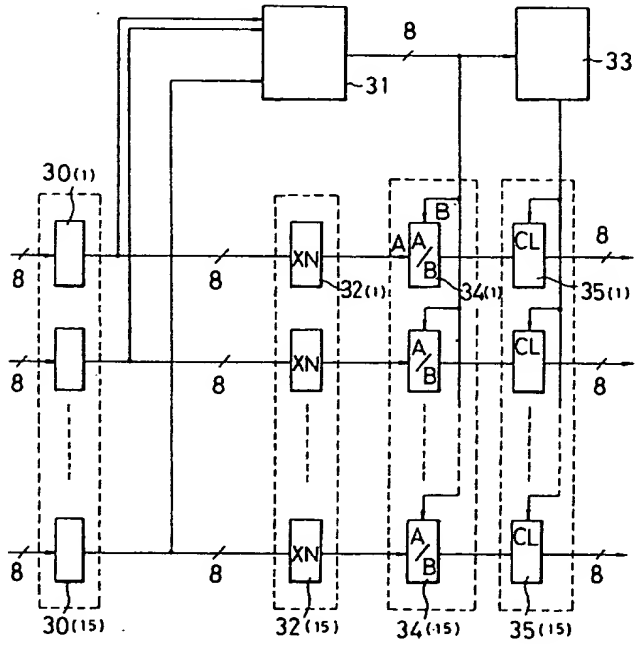
第 1 図


1
ワイヤレスマイク
11
14
マイク
13
プリアンプ
15
プリエンファシス
FM受信機
12
2
16(1)
B.P.F. 1
17(1)
サンプルホールド1
B.P.F. 2
サンプルホールド2
16(2)
16(15)
17(2)
17(15)
B.P.F 15
サンプルホールド15
マルチプレクサ
19
18
8ビットA/D
I/Oポート
20
レベル検出器
25
3
24
CPU
システムプログラム
23
入力パターンメモリ
22
登録パターンメモリ
21
I/Oポート
402
4
403
制御部
401
メモリ部
8ビットD/A
404
ローパスフィルター
405
カラーTV
406

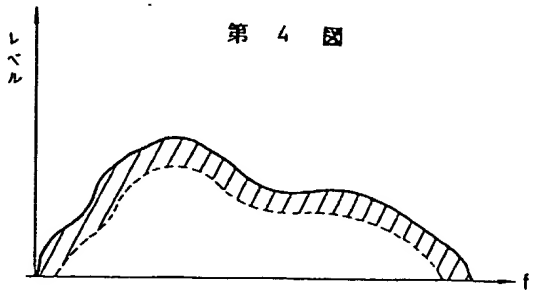
第 2 図


レベル
f (周波数)
f2
f1
0
t1
t2
t3
t(時間)

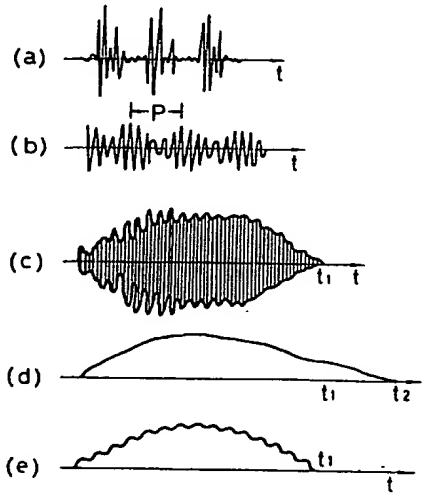
第 3 図


8
31
33
30(1)
8
XN
32(1)
A
B
A/B
34(1)
CL
35(1)
30(15)
32(15)
34(15)
35(15)

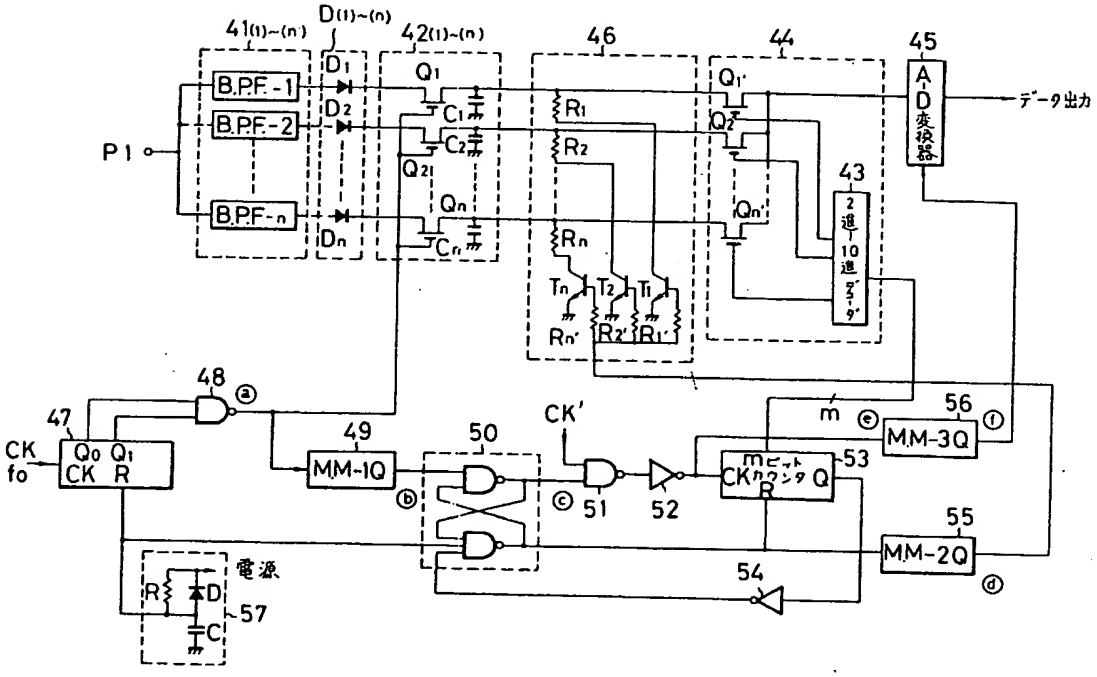
第 4 図


レベル
f

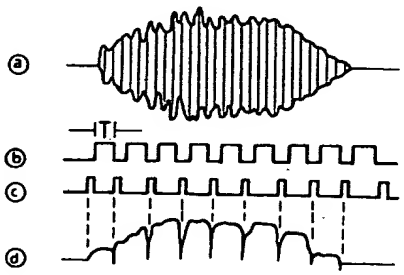
第 5 図


(a)
t
P
(b)
t
(c)
t1
t
(d)
t1
t2
(e)
t1
t

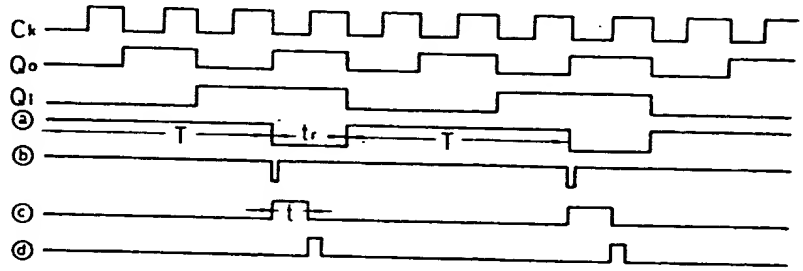
第 6 図


41(1)~(n)
D(1)~(n)
42(1)~(n)
46
44
45
B.P.F.-1
B.P.F.-2
B.P.F.-n
P1
D1
D2
Dn
Q1
C1
C2
Q21
Qn
Cn
R1
R2
Rn
Tn
T2
T1
Rn'
R2'
R1'
Q1'
Q2'
Qn'
A-D変換器
データ出力
43
2進10進データ
48
47
CK
fo
Q0 Q1
CK R
49
MM-1Q
50
CK'
51
52
mビットカウンタ
CK R Q
53
m
56
MM-3Q
55
MM-2Q
54
電源
R
D
C
57

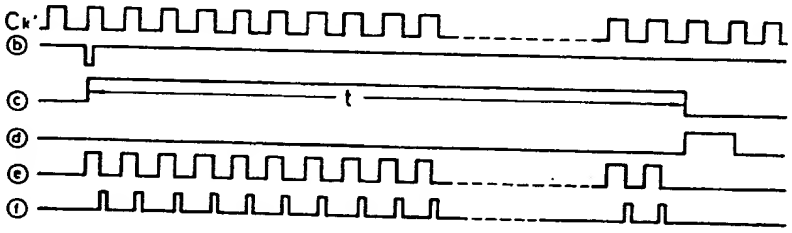
第 7 図



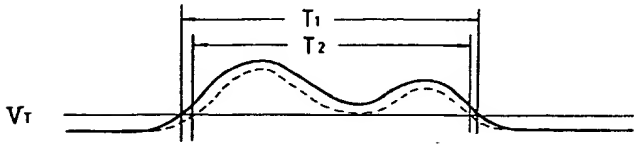
第 8 図



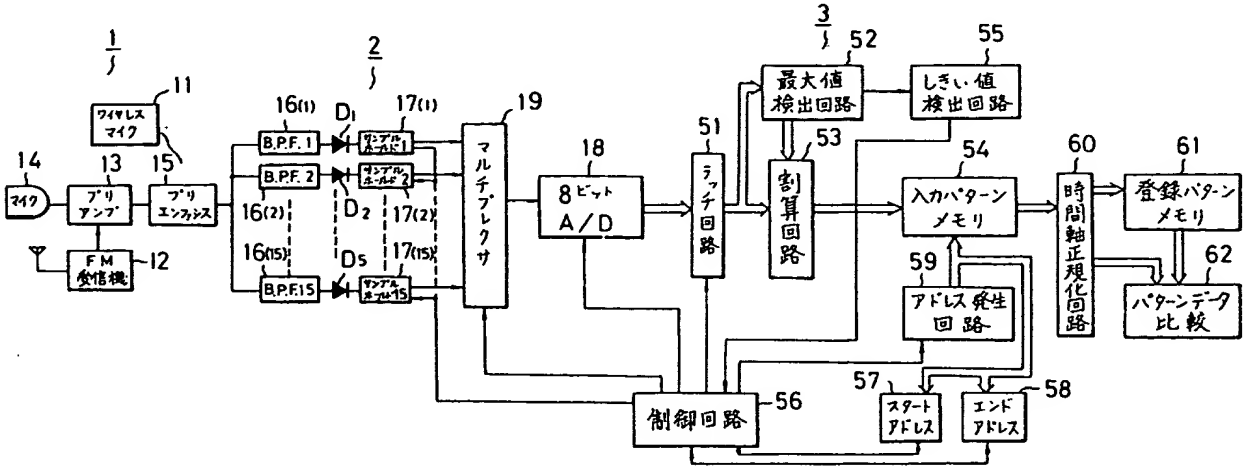
第 9 図



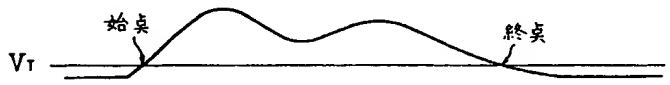
第 10 図



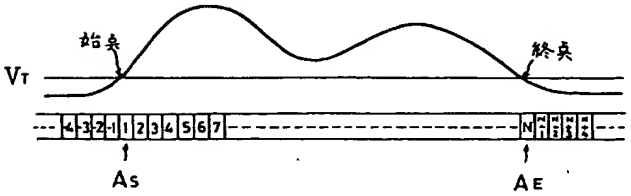
第 11 図



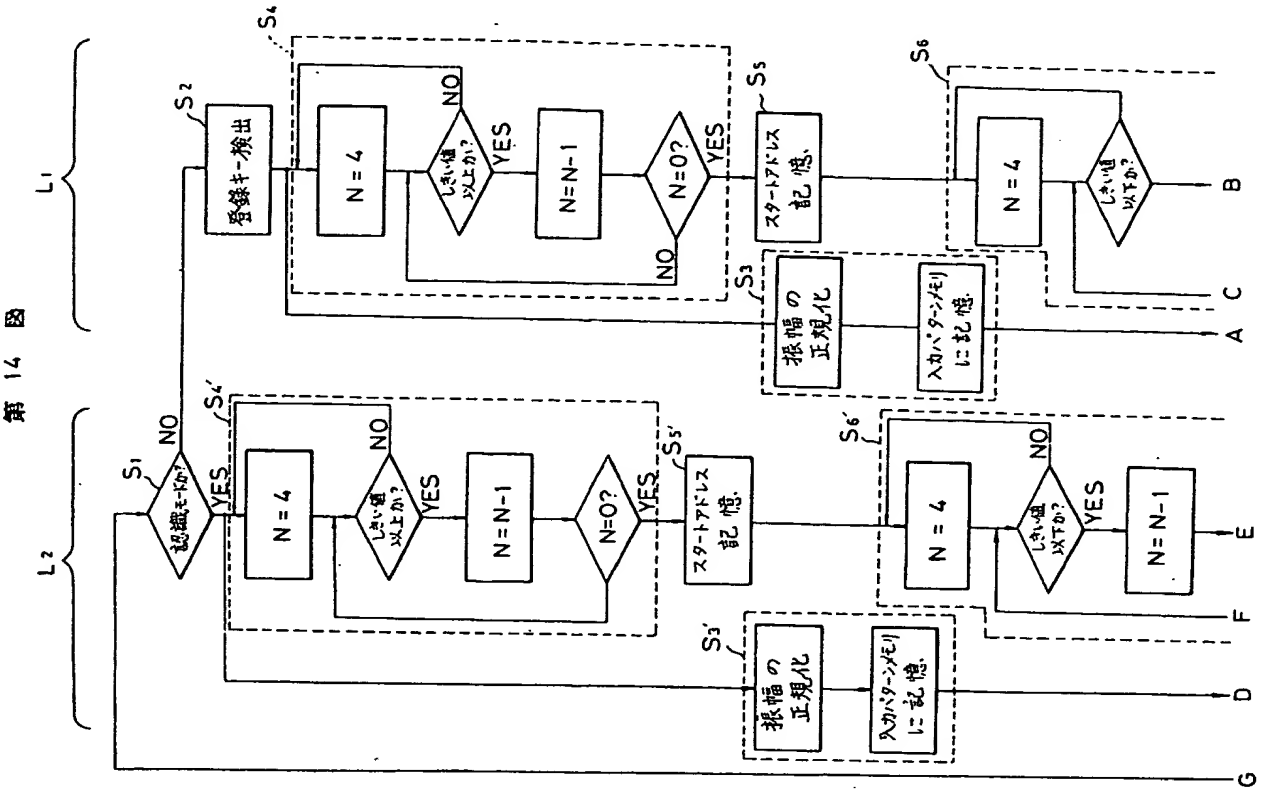
第 12 図



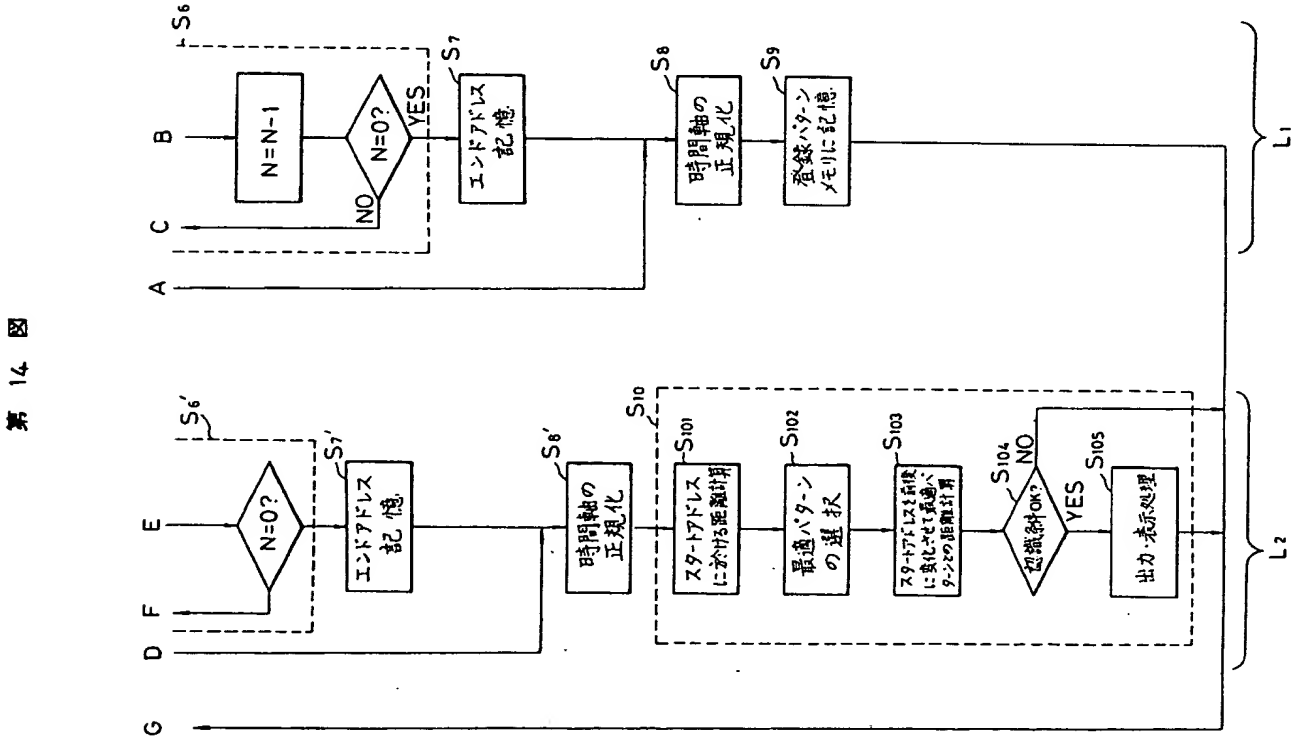
第 13 図



第14図


L1
L2
S1
S2
S3
S4
S5
S6
S1'
S3'
S4'
S5'
S6'
登録キー検出
認識キーON?
N=4
しきい値以上か?
N=N-1
N=0?
スタートアドレス記憶
振幅の正規化
入力パターンメモリに記憶
しきい値以下か?
YES
NO
A
B
C
D
E
F
G

第14図


L1
L2
S6
S7
S8
S9
S10
S6'
S7'
S8'
S101
S102
S103
S104
S105
N=N-1
N=0?
エンドアドレス記憶
時間軸の正規化
登録パターンメモリに記憶
スタートアドレスにおける距離計算
最適パターンの選択
スタートアドレスを順次変化させて最適パターンとの距離計算
認識OK?
出力・表示処理
YES
NO
A
B
C
D
E
F
G

手続補正書（方式）

昭和57年1月　日

特許庁長官　若杉和夫　殿

1. 事件の表示

特願昭57－133573号

2. 発明の名称

音声認識装置

3. 補正をする者

事件との関係　特許出願人

（307）東京芝浦電気株式会社

4. 代理人

住所　東京都港区虎ノ門1丁目26番5号　第17森ビル
〒105　電話 03（502）3181（大代表）

氏名（5847）弁理士 鈴江武彦

5. 補正命令の日付

昭和57年10月26日

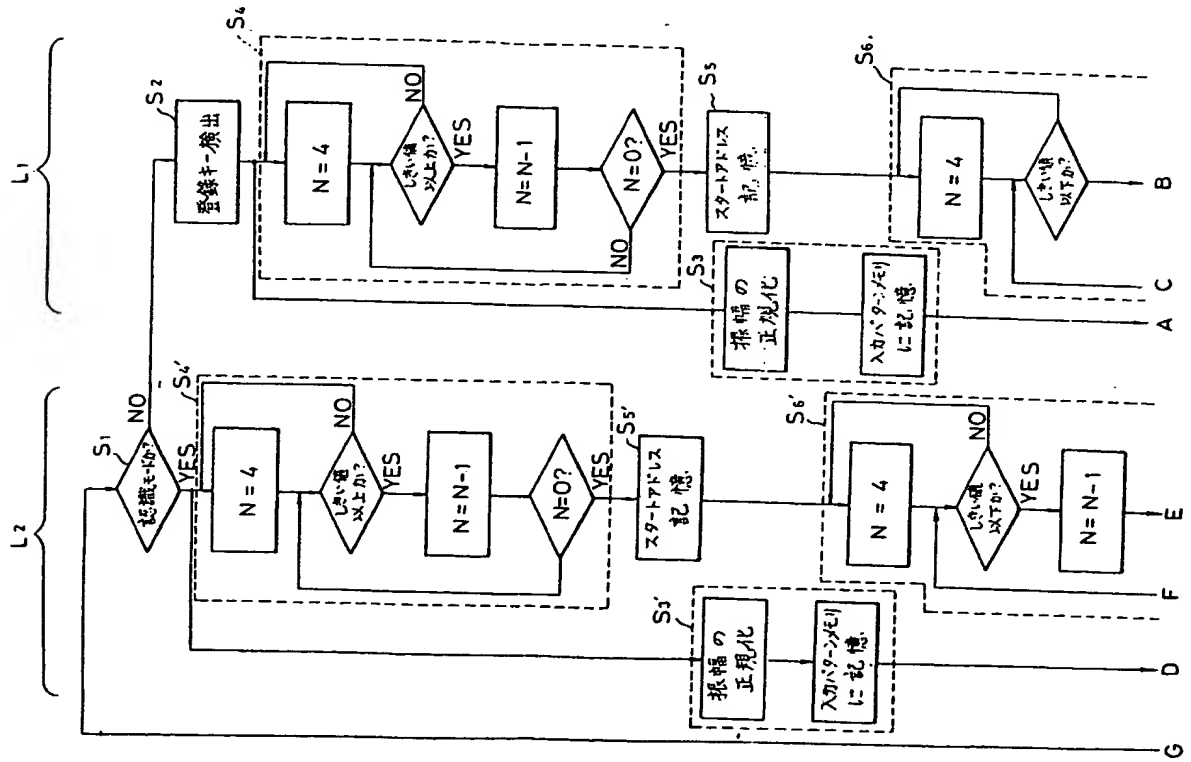
6. 補正の対象

図面

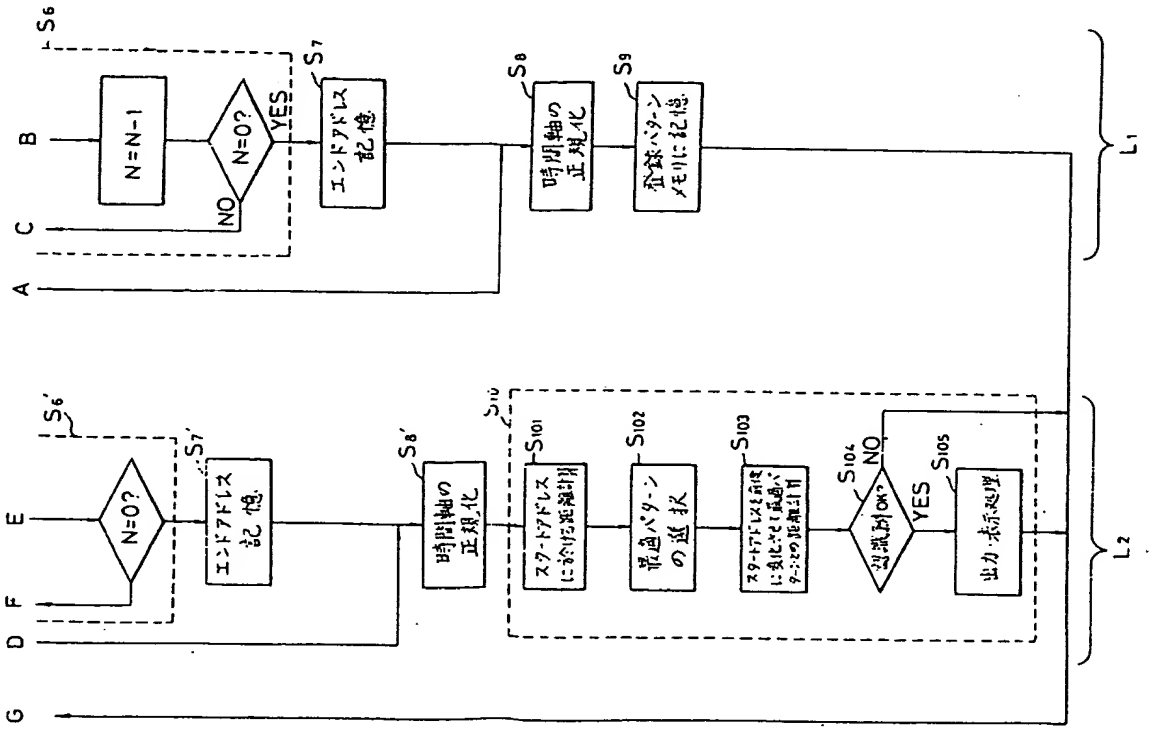
7. 補正の内容

図面第14図の図番を、別紙に朱記の如く「第14図(1)」，「第14図(2)」と訂正する。

第14図(1)



第 14 図 (2)


S6
N=N-1
N=0?
NO
YES
S7
エンドアドレス記憶
S8
時間軸の正規化
S9
登録パターンメモリに記憶
L1
S6'
N=0?
S7'
エンドアドレス記憶
S8'
時間軸の正規化
S10
S101
スタートアドレスに対応する距離計算
S102
最適パターンの選択
S103
S104
NO
YES
S105
出力・表示処理
L2
A
B
C
D
E
F
G